Black-Box Optimization Benchmarking of NEWUOA compared to BIPOP-CMA-ES

Raymond Ros, Nikolaus Hansen

Black-Box Optimization Benchmarking of NEWUOA compared to BIPOP-CMA-ES

On the BBOB Noiseless Testbed

Nikolaus Hansen
Microsoft Research–INRIA Joint Centre
28 rue Jean Rostand
91893 Orsay Cedex, France
Nikolaus.Hansen@inria.fr

Raymond Ros
TAO Team-Project – INRIA Saclay
LRI, Bat 490, Univ. Paris-Sud
F-91405 Orsay Cedex, France
raymond.ros@inria.fr

ABSTRACT

In this paper, the performances of the NEW Unconstrained Optimization Algorithm (NEWUOA) on some noiseless functions are compared to those of the BI-POPulation Covariance Matrix Adaptation-Evolution Strategy (BIPOP-CMA-ES). The two algorithms were benchmarked on the BBOB 2009 noiseless function testbed. The comparison shows that NEWUOA outperforms BIPOP-CMA-ES on some functions like the Sphere or the Rosenbrock functions. Also the independent restart procedure used for NEWUOA allows it to perform better than BIPOP-CMA-ES on the Gallagher functions. Nevertheless, BIPOP-CMA-ES is faster and has a better success probability than NEWUOA in reaching target function values smaller than one on all other functions.

Categories and Subject Descriptors

G.1.6 [**Numerical Analysis**]: Optimization—*global optimization, unconstrained optimization*; F.2.1 [**Analysis of Algorithms and Problem Complexity**]: Numerical Algorithms and Problems

General Terms

Algorithms

Keywords

Benchmarking, Black-box optimization, Evolution strategy, Derivative-free optimization

1. INTRODUCTION

In the context of Black-Box Optimization (BBO), there is the community of Derivative-Free Optimization (DFO) which has proposed the NEW Unconstrained Optimization Algorithm (NEWUOA) [8]. NEWUOA is a trust-region method. NEWUOA computes a quadratic interpolation of the objective function in the current trust region and performs a a truncated conjugate gradient minimization of the surrogate model in the trust region.

From the community of evolutionary computation, the Covariance Matrix Adaptation-Evolution Strategy (CMA-ES) is a state-of-the-art stochastic population-based search method. A BI-POPulation (BIPOP) CMA-ES was introduced in [4] and makes use of a multi-start strategy and two populations with different sizes.

Comparisons of NEWUOA and CMA-ES on a small set of essentially unimodal functions were done in [1, 2]. NEWUOA considerably outperforms CMA-ES on well-conditioned problems and on convex problems with moderate condition number. It performs particularly well on separable convex problems. On non-convex (unimodal) problems with a moderate condition number of 10^4 and on non-separable problems with a condition number of 10^6, the performance of NEWUOA and CMA-ES align. With even larger condition numbers CMA-ES becomes somewhat advantageous.

In this paper, we compare NEWUOA to BIPOP-CMA-ES based on the experimental data obtained for the Black-Box Optimization Benchmarking (BBOB) workshop that was held at the Genetic and Evolutionary Computation COnference 2009.

For more details on the algorithms, their parameter tuning, we refer to [10] for NEWUOA and [4] for BIPOP-CMA-ES.

2. EXPERIMENTAL PROCEDURE

We used the data obtained in [10] for NEWUOA using $2n + 1$ points for interpolating the quadratic model, where n is the dimension of the search space, and the data from [4] for BIPOP-CMA-ES. For benchmarking NEWUOA on the BBOB 2009 noiseless function testbed, an independent multi-start procedure had been implemented as advised in [5]. The BIPOP-CMA-ES includes a restart procedure but adds a population size management policy.

We use the BBOB 2010 post-processing software to compare the performances of NEWUOA and BIPOP-CMA-ES. This can be done without any modifications of the data from BBOB 2009. In any case, the differences in the experimental set-up of BBOB 2009 and BBOB 2010 are minimal. The differences reside in the function instances considered (1 to 5 versus 1 to 15 resp.) and their repetition (3 times versus 1 time resp.).

The parameter settings of NEWUOA and BIPOP-CMA-

GECCO'10, July 7–11, 2010, Portland, Oregon, USA.

ES are described in [10] and [4]. For both algorithms, the crafting effort [6] is equal to CrE= 0.

3. CPU TIMING EXPERIMENTS

For the timing experiments, both algorithms were run on f_8 and restarted until at least 30 seconds (according to [6]). The experiments for NEWUOA has been conducted on a Intel Core 2 6700 processor (2.66 GHz) on Linux 2.6.24.7. The results were 8.1 ; 11 ; 21 ; 58 ; 170 ; 620 and 2500 $\times 10^{-6}$ seconds per function evaluations for NEWUOA in dimensions 2 ; 3 ; 5 ; 10 ; 20 ; 40 and 80 respectively. The results show a dependency between the time per function evaluations and the dimension of the search space.

The experiments for BIPOP-CMA-ES has been conducted on a Intel Core 2 6700 processor (2.66 GHz) on Linux 2.6.24.7 using Matlab R2008a. The results were 6.2 ; 5.8 ; 5.6 ; 5.7 ; 5.8 ; 5.9 and 6.3 $\times 10^{-4}$ seconds per function evaluation for BIPOP-CMA-ES in dimensions 2 ; 3 ; 5 ; 10 ; 20 ; 40 and 80 respectively.

4. RESULTS

Results from experiments according to [6] on the benchmark functions given in [3, 7] are presented in Figures 1, 2, 3 and 4 and in Tables 1 and 2. The **expected running time (ERT)**, used in the figures and table, depends on a given target function value, $f_\mathrm{t} = f_\mathrm{opt} + \Delta f$, and is computed over all relevant trials as the number of function evaluations executed during each trial while the best function value did not reach f_t, summed over all trials and divided by the number of trials that actually reached f_t [6, 9]. **Statistical significance** is tested with the rank-sum test for a given target Δf_t (10^{-8} in Figure 1) using, for each trial, either the number of needed function evaluations to reach Δf_t (inverted and multiplied by -1), or, if the target was not reached, the best Δf-value achieved, measured only up to the smallest number of overall function evaluations for any unsuccessful trial under consideration.

NEWUOA outperforms BIPOP-CMA-ES on f_1 by a factor of about 50 and on the Linear Slope and the Rosenbrock function by a factor of about three. On the other unimodal functions the picture is comparatively mixed, presumably due to local deformations in the function topographies: besides f_1, all functions deviate significantly from a quadratic form. The most surprising results can be observed on the multi-modal functions f_{21} and f_{22}, where NEWUOA consistently outperforms the BIPOP-CMA-ES, for larger dimension and the more difficult target values even by a factor between 10 and 100. The applied independent restarts of NEWUOA appear to be more effective than the large population size of BIPOP-CMA-ES, which is in turn more helpful on the remaining multi-modal functions. On the most difficult multi-modal functions, the performance is not comparable, as BIPOP-CMA-ES were allowed to execute more function evaluations than NEWUOA. Overall, NEWUOA considerably outperforms BIPOP-CMA-ES on about seven functions, while BIPOP-CMA-ES considerably outperforms NEWMAN on about eleven functions.

In conclusion, NEWUOA and BIPOP-CMA-ES are two quite complementary algorithms in their performance. On most problems, one of them considerably outperforms the other. This makes both of them good candidates to be used in an ensemble of black-box search algorithms.

5. REFERENCES

[1] A. Auger, N. Hansen, J. M. Perez Zerpa, R. Ros, and M. Schoenauer. Empirical comparisons of several derivative free optimization algorithms. In *Acte du 9ieme colloque national en calcul des structures*, May 2009.

[2] A. Auger, N. Hansen, J. M. P. Zerpa, R. Ros, and M. Schoenauer. Experimental comparisons of derivative free optimization algorithms. In J. Vahrenhold, editor, *SEA*, volume 5526 of *Lecture Notes in Computer Science*, pages 3–15. Springer, 2009.

[3] S. Finck, N. Hansen, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Presentation of the noiseless functions. Technical Report 2009/20, Research Center PPE, 2009. Updated February 2010.

[4] N. Hansen. Benchmarking a BI-population CMA-ES on the BBOB-2009 function testbed. In Rothlauf [11], pages 2389–2396.

[5] N. Hansen, A. Auger, S. Finck, and R. Ros. Real-parameter black-box optimization benchmarking 2009: Experimental setup. Technical Report RR-6828, INRIA, 2009.

[6] N. Hansen, A. Auger, S. Finck, and R. Ros. Real-parameter black-box optimization benchmarking 2010: Experimental setup. Technical Report RR-7215, INRIA, 2010.

[7] N. Hansen, S. Finck, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Noiseless functions definitions. Technical Report RR-6829, INRIA, 2009. Updated February 2010.

[8] M. J. D. Powell. The NEWUOA software for unconstrained optimization without derivatives. *Large Scale Nonlinear Optimization*, pages 255–297, 2006.

[9] K. Price. Differential evolution vs. the functions of the second ICEO. In *Proceedings of the IEEE International Congress on Evolutionary Computation*, pages 153–157, 1997.

[10] R. Ros. Benchmarking the NEWUOA on the BBOB-2009 function testbed. In Rothlauf [11], pages 2421–2428.

[11] F. Rothlauf, editor. *Genetic and Evolutionary Computation Conference, GECCO 2009, Proceedings, Montreal, Québec, Canada, July 8-12, 2009, Companion Material*. ACM, 2009.

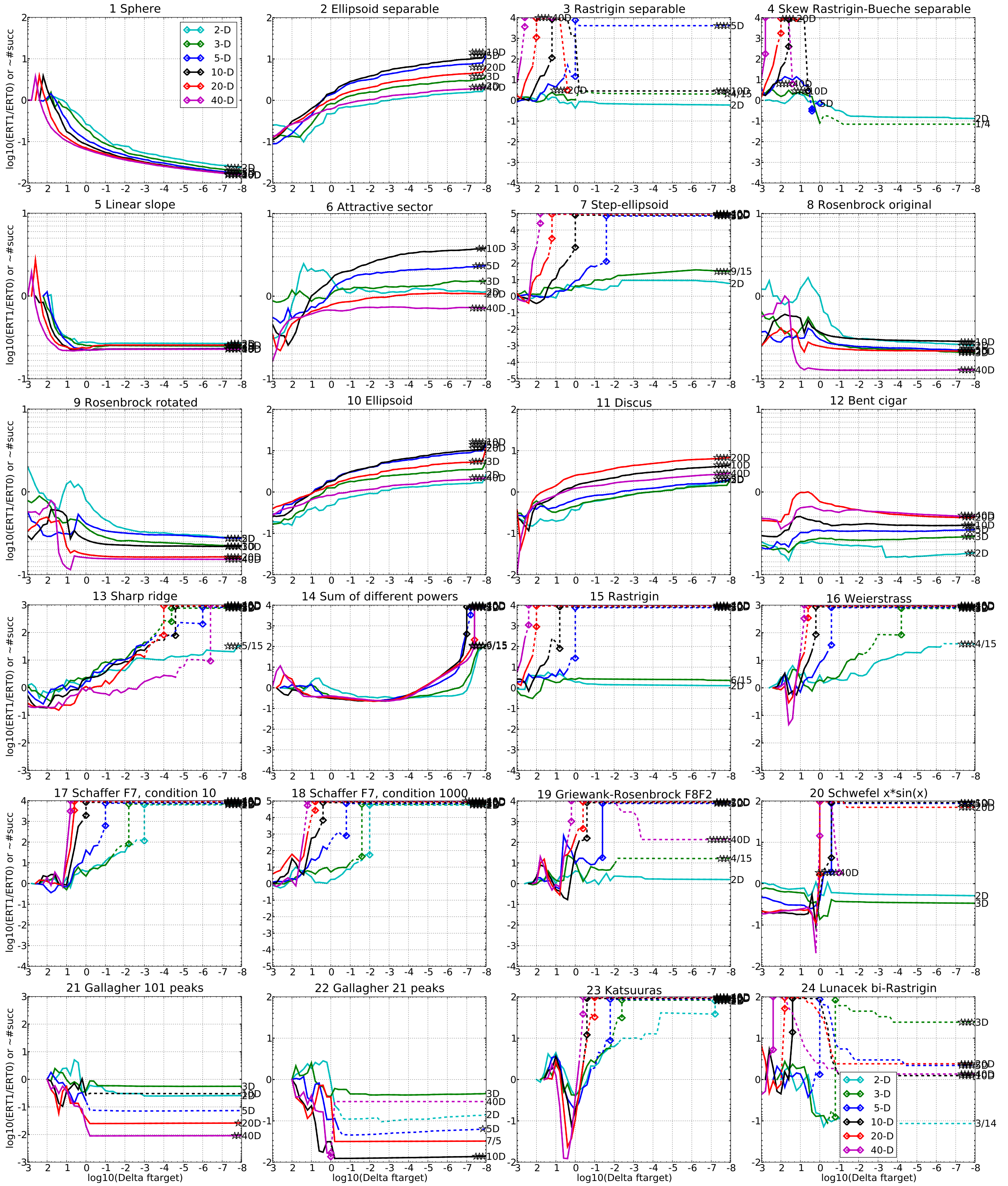

Figure 1: ERT ratio of NEWUOA divided by BIPOP-CMA-ES versus $\log_{10}(\Delta f)$ for f_1–f_{24} in 2, 3, 5, 10, 20, 40-D. Ratios $< 10^0$ indicate an advantage of NEWUOA, smaller values are always better. The line gets dashed when for any algorithm the ERT exceeds thrice the median of the trial-wise overall number of f-evaluations for the same algorithm on this function. Symbols indicate the best achieved Δf-value of one algorithm (ERT gets undefined to the right). The dashed line continues as the fraction of successful trials of the other algorithm, where 0 means 0% and the y-axis limits mean 100%, values below zero for NEWUOA. The line ends when no algorithm reaches Δf anymore. The number of successful trials is given, only if it was in $\{1 \dots 9\}$ for NEWUOA (1st number) and non-zero for BIPOP-CMA-ES (2nd number). Results are significant with $p = 0.05$ for one star and $p = 10^{-\#\star}$ otherwise, with Bonferroni correction within each figure.

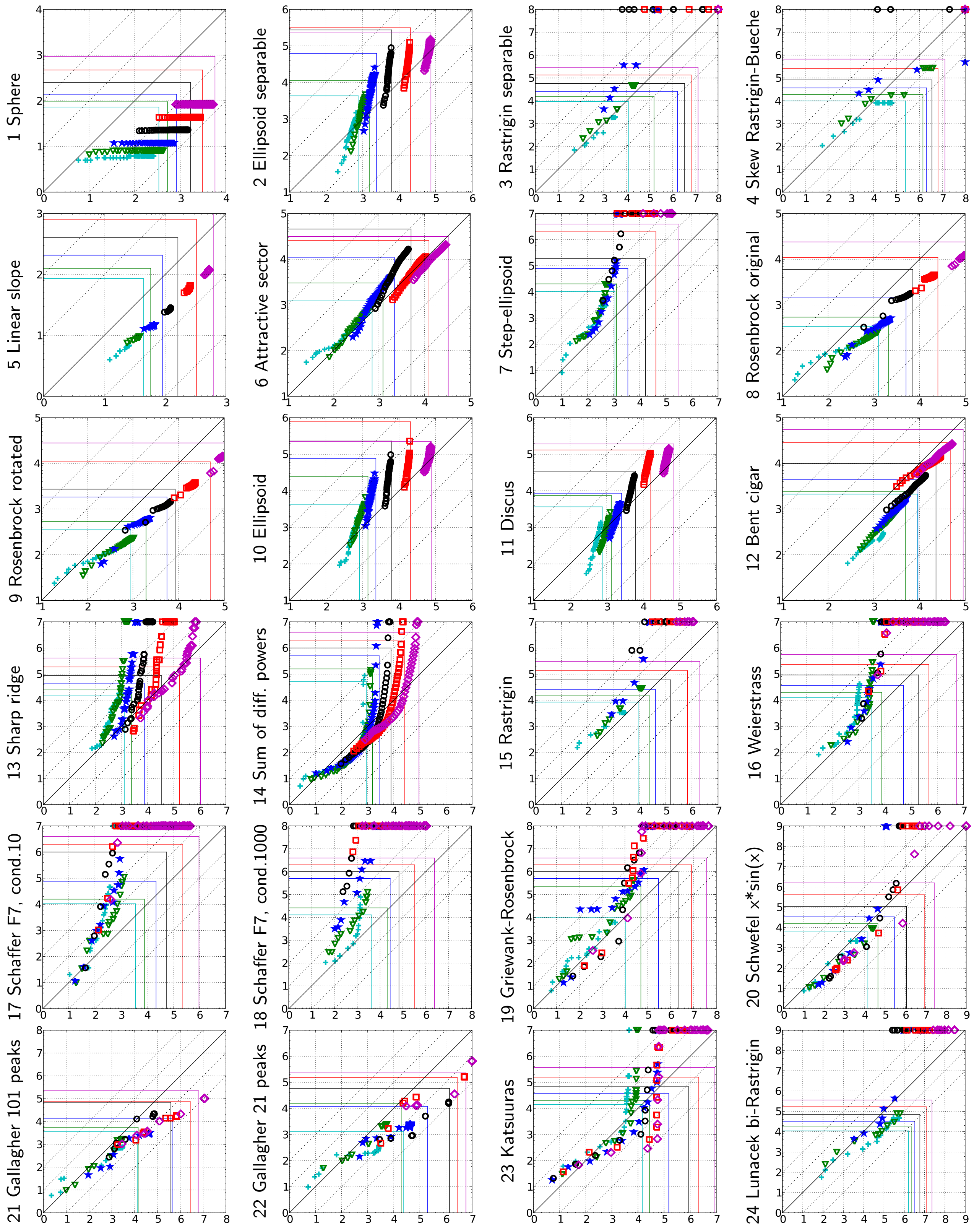

Figure 2: Expected running time (ERT in log10 of number of function evaluations) of NEWUOA versus BIPOP-CMA-ES for 46 target values $\Delta f \in [10^{-8}, 10]$ in each dimension for functions f_1–f_{24}. Markers on the upper or right edge indicate that the target value was never reached by NEWUOA or BIPOP-CMA-ES respectively. Markers represent dimension: 2:+, 3:▽, 5:⋆, 10:○, 20:□, 40:◇. The colored lines indicate maximum number of function evaluations

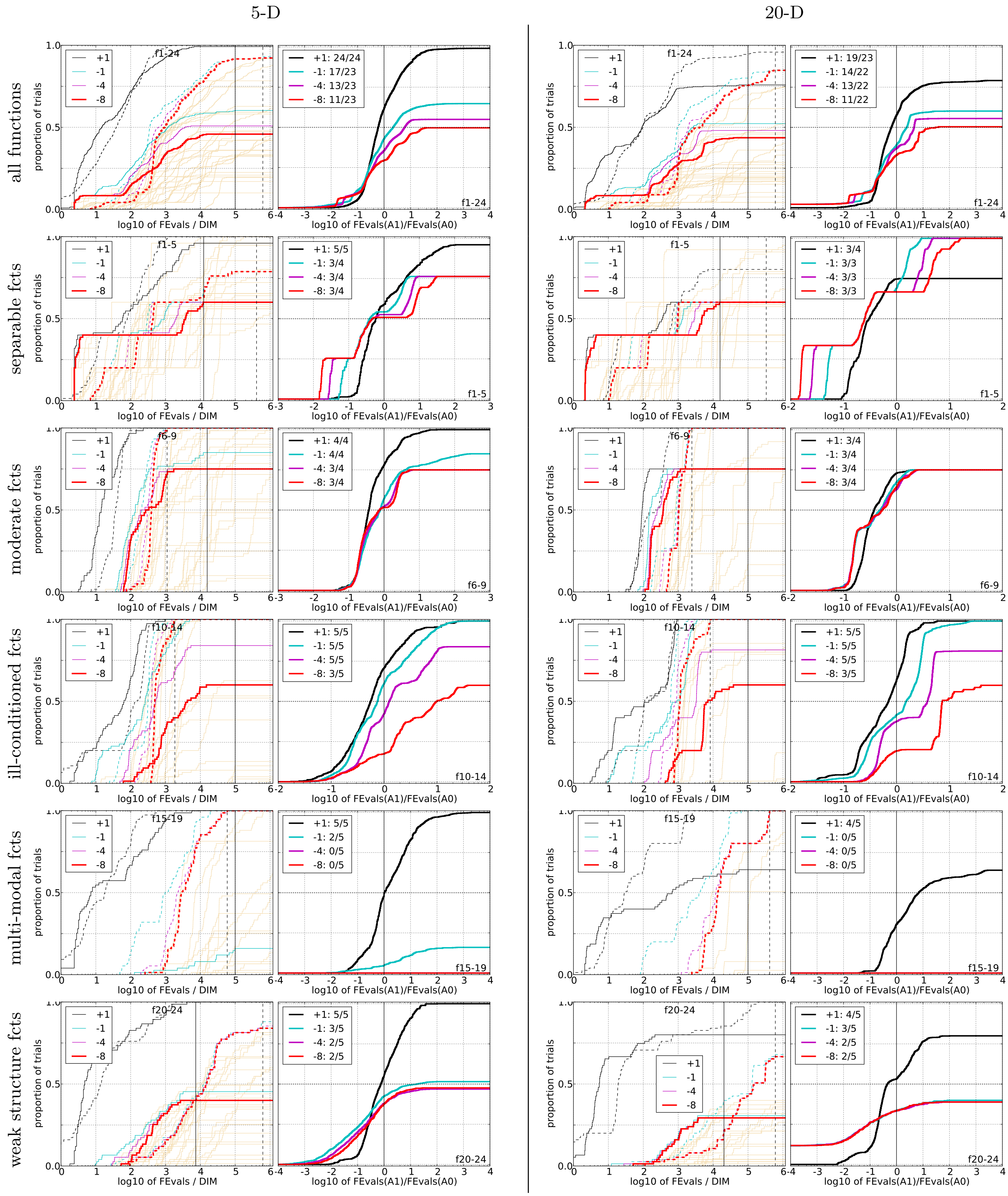

Figure 3: Empirical cumulative distributions (ECDF) of run lengths and speed-up ratios in 5-D (left) and 20-D (right). Left sub-columns: ECDF of the number of function evaluations divided by dimension D (FEvals/D) to reach a target value $f_{\rm opt} + \Delta f$ with $\Delta f = 10^k$, where $k \in \{1, -1, -4, -8\}$ is given by the first value in the legend, for NEWUOA (solid) and BIPOP-CMA-ES (dashed). Light beige lines show the ECDF of FEvals for target value $\Delta f = 10^{-8}$ of algorithms benchmarked during BBOB-2009. Right sub-columns: ECDF of FEval ratios of NEWUOA divided by BIPOP-CMA-ES, all trial pairs for each function. Pairs where both trials failed are disregarded, pairs where one trial failed are visible in the limits being > 0 or < 1. The legends indicate the number of functions that were solved in at least one trial (NEWUOA first).

5-D

Δf	1e+1	1e+0	1e-1	1e-3	1e-5	1e-7	#succ
$\mathbf{f_1}$	11	12	12	12	12	12	15/15
0: BIP	3.2	9.0	15	27	40	53	15/15
1: NEW	1.1	**1**^{*3}	**1**^{*3}	**1**^{*3}	**1**^{*3}	**1**^{*3}	15/15
$\mathbf{f_2}$	83	87	88	90	92	94	15/15
0: BIP	13	16	**18***	**20**^{*2}	**21**^{*3}	**22**^{*3}	15/15
1: NEW	**5.7**^{*2}	22	45	85	129	166	15/15
$\mathbf{f_3}$	716	1622	1637	1646	1650	1654	15/15
0: BIP	1.4	**16**^{*3}	**139**^{*2}	**139**^{*2}	**139**^{*2}	**140**^{*2}	14/15
1: NEW	6.1	229	∞	∞	∞	∞ *2.5e4*	0/15
$\mathbf{f_4}$	809	1633	1688	1817	1886	1903	15/15
0: BIP	**2.7**^{*3}	∞	∞	∞	∞	∞ *1.8e6*	0/15
1: NEW	27	305	∞	∞	∞	∞ *3.4e4*	0/15
$\mathbf{f_5}$	10	10	10	10	10	10	15/15
0: BIP	4.5	6.5	6.6	6.6	6.6	6.6	15/15
1: NEW	**1.3**^{*3}	**1.5**^{*3}	**1.5**^{*3}	**1.5**^{*3}	**1.5**^{*3}	**1.5**^{*3}	15/15
$\mathbf{f_6}$	114	214	281	580	1038	1332	15/15
0: BIP	2.3	2.1	2.2	**1.7***	**1.3**^{*2}	**1.3**^{*2}	15/15
1: NEW	1.7	2.4	3.6	3.3	2.7	2.9	15/15
$\mathbf{f_7}$	24	324	1171	1572	1572	1597	15/15
0: BIP	5.0	1.5	**1**^{*3}	**1**^{*3}	**1**^{*3}	**1**^{*3}	15/15
1: NEW	10	13	60	∞	∞	∞ *2.9e4*	0/15
$\mathbf{f_8}$	73	273	336	391	410	422	15/15
0: BIP	3.2	3.7	4.5	4.8	5.1	5.4	15/15
1: NEW	**1**^{*2}	**1.1**^{*2}	**1.2**^{*3}	**1.2**^{*3}	**1.2**^{*3}	**1.2**^{*3}	15/15
$\mathbf{f_9}$	35	127	214	300	335	369	15/15
0: BIP	5.8	8.7	7.2	6.4	6.3	6.2	15/15
1: NEW	**1.8**^{*3}	3.6	**2.5**^{*2}	**1.9**^{*2}	**1.9**^{*3}	**1.7**^{*3}	15/15
$\mathbf{f_{10}}$	349	500	574	626	829	880	15/15
0: BIP	3.5	2.9	2.7	**2.8**^{*3}	**2.3**^{*3}	**2.4**^{*3}	15/15
1: NEW	3.1	5.5	8.1	14	16	21	15/15
$\mathbf{f_{11}}$	143	202	763	1177	1467	1673	15/15
0: BIP	8.4	7.2	2.2	1.6	**1.4**^{*3}	**1.3**^{*3}	15/15
1: NEW	**3.5**^{*3}	**4.7***	1.8	1.8	2.0	2.2	15/15
$\mathbf{f_{12}}$	108	268	371	461	1303	1494	15/15
0: BIP	11	7.4	7.4	7.7	3.3	3.3	15/15
1: NEW	3.5	**2.6***	**2.5***	**2.6**^{*2}	**1.1**^{*2}	**1.1***	15/15
$\mathbf{f_{13}}$	132	195	250	1310	1752	2255	15/15
0: BIP	3.9	5.4	5.9	**1.6**^{*3}	**1.5**^{*3}	**1.7**^{*3}	15/15
1: NEW	3.1	9.3	35	54	335	∞ *4.0e4*	0/15
$\mathbf{f_{14}}$	10	41	58	139	251	476	15/15
0: BIP	1.1	2.8	3.7	4.6	5.4	**4.5**^{*3}	15/15
1: NEW	1.7	**1**^{*3}	**1**^{*3}	**1.2**^{*3}	5.5	2525	0/15
$\mathbf{f_{15}}$	511	9310	19369	20073	20769	21359	14/15
0: BIP	1.6	**1.5**^{*3}	**1.2**^{*2}	**1.2**^{*2}	**1.2**^{*2}	**1.2**^{*2}	15/15
1: NEW	5.8	41	∞	∞	∞	∞ *2.5e4*	0/15
$\mathbf{f_{16}}$	120	612	2662	10449	11644	12095	15/15
0: BIP	3.0	**3.6**^{*2}	**2.6**^{*3}	**1.3**^{*3}	**1.4**^{*3}	**1.4**^{*3}	15/15
1: NEW	2.1	29	∞	∞	∞	∞ *3.6e4*	0/15
$\mathbf{f_{17}}$	5.2	215	899	3669	6351	7934	15/15
0: BIP	3.4	**1***	**1**^{*3}	**1**^{*3}	**1**^{*3}	**1.2**^{*3}	15/15
1: NEW	2.3	40	617	∞	∞	∞ *3.4e4*	0/15
$\mathbf{f_{18}}$	103	378	3968	9280	10905	12469	15/15
0: BIP	1	**3.4**^{*3}	**1**^{*3}	**1**^{*3}	**1.2**^{*3}	**1.3**^{*3}	15/15
1: NEW	31	1351	∞	∞	∞	∞ *9.2e4*	0/15
$\mathbf{f_{19}}$	1	1	242	1.20e5	1.21e5	1.22e5	15/15
0: BIP	20	2801	161	**1**^{*3}	**1**^{*3}	**1**^{*3}	15/15
1: NEW	14	26728	1415	∞	∞	∞ *5.0e5*	0/15
$\mathbf{f_{20}}$	16	851	38111	54470	54861	55313	14/15
0: BIP	3.3	8.2	2.8	2.1	2.2	2.2	15/15
1: NEW	**1**^{*2}	3.3	∞	∞	∞	∞ *3.2e4*	0/15
$\mathbf{f_{21}}$	41	1157	1674	1705	1729	1757	14/15
0: BIP	2.3	14	24	25	25	25	15/15
1: NEW	1.1	2.2	1.8	1.8	1.8	1.9	15/15
$\mathbf{f_{22}}$	71	386	938	1008	1040	1068	14/15
0: BIP	6.9	20	45	42	41	40	15/15
1: NEW	2.1	2.1	2.0	2.1	2.3	2.4	15/15
$\mathbf{f_{23}}$	3.0	518	14249	31654	33030	34256	15/15
0: BIP	**1.7**^{*2}	13	3.7	1.8	1.8	1.8	15/15
1: NEW	6.2	2.4	7.1	∞	∞	∞ *3.3e4*	0/15
$\mathbf{f_{24}}$	1622	2.16e5	6.36e6	9.62e6	1.28e7	1.28e7	3/15
0: BIP	2.1	1.6	1	1	1	1	3/15
1: NEW	2.9	2.1	∞	∞	∞	∞ *3.0e4*	0/15

20-D

Δf	1e+1	1e+0	1e-1	1e-3	1e-5	1e-7	#succ
$\mathbf{f_1}$	43	43	43	43	43	43	15/15
0: BIP	7.9	14	20	33	45	57	15/15
1: NEW	**1.0**^{*3}	**1.0**^{*3}	**1.0**^{*3}	**1.0**^{*3}	**1.0**^{*3}	**1.0**^{*3}	15/15
$\mathbf{f_2}$	385	386	387	390	391	393	15/15
0: BIP	35	40	**44**^{*2}	**47**^{*3}	**48**^{*3}	**50**^{*3}	15/15
1: NEW	**18**^{*3}	42	71	125	174	219	15/15
$\mathbf{f_3}$	5066	7626	7635	7643	7646	7651	15/15
0: BIP	**12**^{*3}	∞	∞	∞	∞	∞ *5.7e6*	0/15
1: NEW	∞	∞	∞	∞	∞	∞ *1.3e5*	0/15
$\mathbf{f_4}$	4722	7628	7666	7700	7758	1.41e5	9/15
0: BIP	∞	∞	∞	∞	∞	∞ *5.5e6*	0/15
1: NEW	∞	∞	∞	∞	∞	∞ *2.2e5*	0/15
$\mathbf{f_5}$	41	41	41	41	41	41	15/15
0: BIP	5.1	6.2	6.3	6.3	6.3	6.3	15/15
1: NEW	**1.2**^{*3}	**1.5**^{*3}	**1.6**^{*3}	**1.6**^{*3}	**1.6**^{*3}	**1.6**^{*3}	15/15
$\mathbf{f_6}$	1296	2343	3413	5220	6728	8409	15/15
0: BIP	1.5	1.3	1.2	1.1	1.2	1.2	15/15
1: NEW	**1**^{*2}	1	1	1.1	1.3	1.3	15/15
$\mathbf{f_7}$	1351	4274	9503	16524	16524	16969	15/15
0: BIP	**1**^{*3}	**4.9**^{*3}	**3.5**^{*3}	**2.2**^{*3}	**2.2**^{*3}	**2.1**^{*3}	15/15
1: NEW	∞	∞	∞	∞	∞	∞ *4.8e5*	0/15
$\mathbf{f_8}$	2039	3871	4040	4219	4371	4484	15/15
0: BIP	4.0	4.0	4.3	4.5	4.6	4.6	15/15
1: NEW	**1**^{*3}	**1**^{*3}	**1**^{*3}	**1**^{*3}	**1**^{*3}	**1**^{*3}	15/15
$\mathbf{f_9}$	1716	3102	3277	3455	3594	3727	15/15
0: BIP	4.7	5.7	6.0	6.1	6.1	6.1	15/15
1: NEW	**1.0**^{*3}	**1**^{*3}	**1**^{*3}	**1**^{*3}	**1**^{*3}	**1**^{*3}	15/15
$\mathbf{f_{10}}$	7413	8661	10735	14920	17073	17476	15/15
0: BIP	1.9	**1.8**^{*2}	**1.6**^{*3}	**1.2**^{*3}	**1.1**^{*3}	**1.1**^{*3}	15/15
1: NEW	1.7	2.6	3.3	4.0	4.7	5.8	15/15
$\mathbf{f_{11}}$	1002	2228	6278	9762	12285	14831	15/15
0: BIP	**10**^{*3}	**5.1**^{*3}	**1.9**^{*3}	**1.4**^{*3}	**1.2**^{*3}	**1.0**^{*3}	15/15
1: NEW	15	13	5.8	6.1	6.6	6.5	15/15
$\mathbf{f_{12}}$	1042	1938	2740	4140	12407	13827	15/15
0: BIP	3.0	4.0	4.5	4.5	1.9	2.0	15/15
1: NEW	3.0	3.0	3.0	2.5	**1**^{*2}	**1**^{*3}	15/15
$\mathbf{f_{13}}$	652	2021	2751	18749	24455	30201	15/15
0: BIP	4.3	2.7	5.1	**1.5**^{*2}	**2.3**^{*3}	**3.0**^{*3}	15/15
1: NEW	**1***	3.0	9.3	19	∞	∞ *1.8e5*	0/15
$\mathbf{f_{14}}$	75	239	304	932	1648	15661	15/15
0: BIP	3.9	2.9	3.7	4.1	**6.2**^{*3}	**1.2**^{*3}	15/15
1: NEW	**1.5**^{*3}	**1**^{*3}	**1**^{*3}	**1**^{*3}	9.1	43	0/15
$\mathbf{f_{15}}$	30378	1.47e5	3.12e5	3.20e5	4.49e5	4.59e5	15/15
0: BIP	**1**^{*3}	**2.0**^{*3}	**1.4**^{*3}	**1.4**^{*3}	**1**^{*3}	**1**^{*3}	15/15
1: NEW	∞	∞	∞	∞	∞	∞ *1.3e5*	0/15
$\mathbf{f_{16}}$	1384	27265	77015	1.88e5	1.98e5	2.20e5	15/15
0: BIP	**1.7**^{*2}	**1.0**^{*3}	**1.2**^{*3}	**1**^{*3}	**1**^{*3}	**1**^{*3}	15/15
1: NEW	16	∞	∞	∞	∞	∞ *2.3e5*	0/15
$\mathbf{f_{17}}$	63	1030	4005	30677	56288	80472	15/15
0: BIP	2.2	**1**^{*3}	**1**^{*3}	**1.2**^{*3}	**1.3**^{*3}	**1.4**^{*3}	15/15
1: NEW	16	∞	∞	∞	∞	∞ *1.5e6*	0/15
$\mathbf{f_{18}}$	621	3972	19561	67569	1.31e5	1.47e5	15/15
0: BIP	**1.0**^{*3}	**2.4**^{*3}	**1.2**^{*3}	**1.1**^{*3}	**1.7**^{*3}	**1.6**^{*3}	15/15
1: NEW	11930	∞	∞	∞	∞	∞ *1.6e6*	0/15
$\mathbf{f_{19}}$	1	1	3.43e5	6.22e6	6.69e6	6.74e6	15/15
0: BIP	169	**23770***	**1.2**^{*3}	**1**^{*3}	**1**^{*3}	**1**^{*3}	15/15
1: NEW	**76**^{*2}	4.29e6	∞	∞	∞	∞ *2.0e6*	0/15
$\mathbf{f_{20}}$	82	46150	3.10e6	5.54e6	5.59e6	5.64e6	14/15
0: BIP	4.3	9.2	1	1	1	1	14/15
1: NEW	**1**^{*3}	15	∞	∞	∞	∞ *3.8e5*	0/15
$\mathbf{f_{21}}$	561	6541	14103	14643	15567	17589	15/15
0: BIP	3.2	55	48	46	43	39	13/15
1: NEW	1.7	2.2	1.2	1.2	1.1	1	15/15
$\mathbf{f_{22}}$	467	5580	23491	24948	26847	1.35e5	12/15
0: BIP	6.8	13	215	202	188	37	5/15
1: NEW	**1***	4.9	6.8	6.4	6.0	1.2	7/15
$\mathbf{f_{23}}$	3.2	1614	67457	4.89e5	8.11e5	8.38e5	15/15
0: BIP	4.3	32	**1**^{*3}	**2.0**^{*2}	**1.2**^{*2}	**1.2**^{*2}	15/15
1: NEW	12	**3.5**^{*3}	32	∞	∞	∞ *1.5e5*	0/15
$\mathbf{f_{24}}$	1.34e6	7.48e6	5.19e7	5.20e7	5.20e7	5.20e7	3/15
0: BIP	**1**^{*2}	**1**^{*2}	**1**^{*2}	**1**^{*2}	**1**^{*2}	**1**^{*2}	3/15
1: NEW	∞	∞	∞	∞	∞	∞ *1.7e5*	0/15

Table 1: Expected running time (ERT in number of function evaluations) divided by the best ERT measured during BBOB-2009 (given in the respective first row) for different Δf values for functions f_1–f_{24}. The median number of conducted function evaluations is additionally given in *italics*, if $\mathrm{ERT}(10^{-7}) = \infty$. #succ is the number of trials that reached the final target $f_{\mathrm{opt}} + 10^{-8}$. 0: BIP is BIPOP-CMA-ES and 1: NEW is NEWUOA. Bold entries are statistically significantly better compared to the other algorithm, with $p = 0.05$ or $p = 10^{-k}$ where $k > 1$ is the number following the $\star$ symbol, with Bonferroni correction of 48.

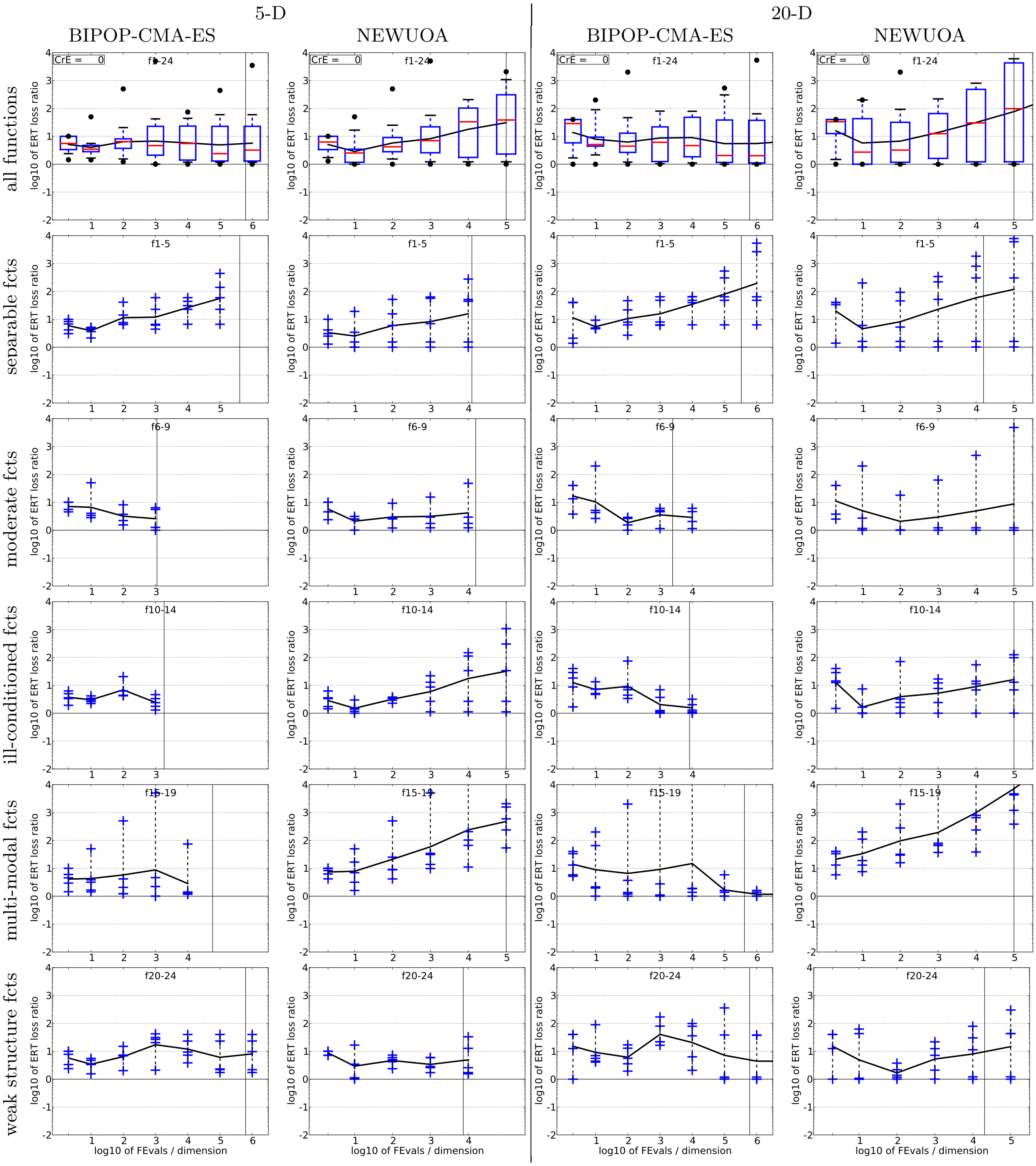

Figure 4: ERT loss ratio versus given budget FEvals. The target value f_{t} for ERT is the smallest (best) recorded function value such that $\mathrm{ERT}(f_{\mathrm{t}}) \leq \mathrm{FEvals}$ for the presented algorithm. Shown is FEvals divided by the respective best $\mathrm{ERT}(f_{\mathrm{t}})$ from BBOB-2009 for functions f_1–f_{24} in 5-D and 20-D. Each ERT is multiplied by $\exp(\mathrm{CrE})$ correcting for the parameter crafting effort. Line: geometric mean. Box-Whisker error bar: 25-75%-ile with median (box), 10-90%-ile (caps), and minimum and maximum ERT loss ratio (points). The vertical line gives the maximal number of function evaluations in this function subset.

Table 2: ERT **loss ratio (see Figure 4) compared to the respective best result from BBOB-2009 for budgets given in the first column. The last row $\mathrm{RL_{US}/D}$ gives the number of function evaluations in unsuccessful runs divided by dimension. Shown are the smallest, 10%-ile, 25%-ile, 50%-ile, 75%-ile and 90%-ile value (smaller values are better). ERT Loss ratio is equal to zero if the algorithm considered outperformed all algorithms from BBOB-2009.**

BIPOP-CMA-ES

	f_1–f_{24} **in 5-D**, maxFE/D=622854					
#FEs/D	best	10%	25%	**med**	75%	90%
2	1.4	2.3	3.3	5.3	9.2	10
10	1.4	1.6	2.7	3.4	4.6	10
100	1.2	1.5	3.0	6.4	7.9	23
1e3	1.0	1.0	1.9	4.6	22	44
1e4	1.0	1.2	1.4	5.1	23	46
1e5	1.0	1.2	1.3	2.3	15	68
1e6	1.0	1.2	1.3	2.8	16	68
$\mathrm{RL_{US}/D}$	3e5	3e5	4e5	4e5	6e5	6e5

NEWUOA

	f_1–f_{24} **in 5-D**, maxFE/D=100000					
#FEs/D	best	10%	25%	**med**	75%	90%
2	1.3	1.7	3.2	6.2	10	10
10	1.0	1.0	1.1	2.5	3.4	17
100	1.0	1.5	2.7	4.2	8.9	28
1e3	1.0	1.2	2.2	6.5	19	57
1e4	1.0	1.2	1.7	23	84	2.1e2
1e5	1.0	1.2	2.0	36	3.1e2	1.1e3
$\mathrm{RL_{US}/D}$	4e3	5e3	6e3	7e3	1e4	1e5

BIPOP-CMA-ES

	f_1–f_{24} **in 20-D**, maxFE/D=605134					
#FEs/D	best	10%	25%	**med**	75%	90%
2	1.0	1.7	5.5	23	40	40
10	1.0	2.1	4.4	5.0	8.3	1.0e2
100	1.0	1.2	2.3	4.1	11	49
1e3	1.0	1.0	1.2	6.1	22	89
1e4	1.0	1.1	1.6	3.9	44	81
1e5	1.0	1.0	1.1	2.0	22	3.1e2
1e6	1.0	1.0	1.1	1.8	22	3.2e2
1e7	1.0	1.0	1.1	1.8	22	2.7e3
$\mathrm{RL_{US}/D}$	1e5	1e5	3e5	3e5	3e5	5e5

NEWUOA

	f_1–f_{24} **in 20-D**, maxFE/D=100000					
#FEs/D	best	10%	25%	**med**	75%	90%
2	1.0	1.5	9.2	33	40	40
10	1.0	1.0	1.0	2.2	31	2.0e2
100	1.0	1.0	1.1	2.8	31	1.1e2
1e3	1.0	1.0	1.4	12	64	2.3e2
1e4	1.0	1.0	1.2	22	3.9e2	9.0e2
1e5	1.0	1.0	1.2	71	2.7e3	6.2e3
1e6	1.0	1.0	1.2	3.7e2	1.7e4	4.6e4
$\mathrm{RL_{US}/D}$	5e3	6e3	8e3	1e4	7e4	1e5